Assignment 1 - CSC/DSC 462 - FALL 2018 - Due September 20

All questions are worth equal marks.

Q1: The Inclusion-Exclusion Principle. If events $A_1, \ldots, A_n$ are mutually exclusive, the probability of the union is

$$P(\cup_{i=1}^n A_i) = P(A_1) + \ldots + P(A_n).$$

If they are not mutually exclusive, then calculation of the probability of their union can become quite complex. For two events, we have

$$P(A_1 \cup A_2) = P(A_1) + P(A_2) - P(A_1 A_2).$$

This can be extended to n events using the *inclusion-exclusion identity*

$$\begin{aligned} P(\cup_{i=1}^n A_i) &= \sum_i P(A_i) \\ &\quad - \sum_{i<j} P(A_i A_j) \\ &\quad + \sum_{i<j<k} P(A_i A_j A_k) \\ &\quad \vdots \\ &\quad -1^{n+1} P(A_1 A_2 \ldots A_n). \end{aligned}$$

(a) Write explicitly the *inclusion-exclusion identity* for $n = 3$.

(b) Suppose any integer from 1 to 105 inclusive is chosen at random with equal probability, which will be denoted N. What is the probability that N is divisible by at least one of 2, 9 or 13?

Q2: Ten birds are arranged in a row on a powerline. Four are red, six are blue. Suppose the ordering is random, that is, all permutations of the ordering of the ten birds are equally likely.

(a) What is the probability that the four red birds are all adjacent?

(b) What is the probability that no two red birds are adjacent?

HINT: In a problem like this, it can be useful to assign labels to the objects to make them completely distinct. So, the blue birds will be labeled $B_1, \ldots, B_6$, the red birds will be labeled $R_1 \ldots, R_4$. The events in question do not depend on this labeling, so this won't affect the answer.

Next, suppose the blue birds are ordered $B_1, B_2, B_3, B_4, B_5, B_6$. We will make 'slots' on each side of the blue birds:

$$\_\_ B_1 \_\_ B_2 \_\_ B_3 \_\_ B_4 \_\_ B_5 \_\_ B_6 \_\_$$

Thus, to construct an ordering, we first order the blue birds, then place each red bird in one of the 7 slots.

Q3: A standard 52 card playing deck assigns a unique combination of 13 ranks (2,3,4,5,6,7,8,9,10,J,Q,K,A) and 4 suits (Clubs, Diamonds, Hearts, Spades) to each card ($13 \times 4 = 52$). Suppose 5 cards are selected at random. Derive the probability that each of the following events occurs.

(a) All cards are face cards (J,Q,K) of a single color (all black [spades or clubs] or all red [hearts or diamonds]).

(b) All suits are represented at least once.

(c) All ranks are distinct, and of a single color.

Carefully list the *tasks* used in the application of the *rule of product*.

HINT: A number of examples of this type of problem, with solutions, can be found in Practice Problem Set 1.

Q4: In genetics, a genotype consists of two genes, each of which is one of (possibly) several types of alleles. Suppose we consider only two alleles, `r` and `R`. Furthermore, suppose the allele `r` exists in a population with frequency q. Under Hardy-Weinberg equilibrium, genotypes are essentially random samples of 2 alleles, one sampled from each parent. Since the genes are usually not ordered (because we don't know which is maternal and which is paternal), the probability of each possible genotype in an individual is

$$\begin{aligned} P(\texttt{rr}) &= q^2, \\ P(\texttt{rR}) &= 2(1-q)q, \\ P(\texttt{RR}) &= (1-q)^2. \end{aligned}$$

We then note that genotypes of unrelated individuals are independent, but genotypes of related individuals are not independent. For example, when two organisms mate, each passes one allele, selected at random, to the offspring, forming that offspring's genotype (this predicts Mendels' Law of Inheritance). If G_o is an offspring genotype then

$$\begin{aligned} P(G_o = \texttt{rr}) &= q^2, \text{ but} \\ P(G_o = \texttt{rr} \mid \text{ both parents have genotype } \texttt{rR}) &= 1/4. \end{aligned}$$

Now, suppose `r` is a *recessive* allele, meaning that it only determines a trait when the genotype is `rr`. Such a trait is a *recessive trait*. Typically, a genetic disease is a recessive trait, and `r` is a *rare allele*, meaning q is very small.

Next, suppose we may determine without error whether or not an individual possesses a recessive trait (and therefore has genotype `rr`). Let G_m, G_f, G_1, G_2 be the genotypes of a mother, father and two offspring (ie. siblings). Assume the parental genotypes are independent (ie. the parents are unrelated). Define events

$$\begin{aligned} R_m &= \{ \text{ Mother has recessive trait } \}, \\ R_f &= \{ \text{ Father has recessive trait } \}, \\ R_1 &= \{ \text{ Offspring 1 has recessive trait } \}, \\ R_2 &= \{ \text{ Offspring 2 has recessive trait } \}. \end{aligned}$$

(a) Determine the conditional probability $P(R_1 \mid R_m)$, the probability that an offspring possesses the recessive trait given that the mother does.

(b) Determine the conditional probability $P(R_1 \mid R_2)$, the probability that an offspring possesses the recessive trait given that a sibling does.

(c) Give a lower bound for the conditional probabilities of Parts (a) and (b). In other words, to what value does each conditional probability approach as q approaches zero?

Q5: The formal axioms of probability can be stated as follows. Let the set S be the sample space. Let $\mathcal{F}$ be some collection of subsets of S. Not all subsets of S need be in $\mathcal{F}$, but S and $\emptyset$ must be. We assign a probability $P(E)$ to each $E \in \mathcal{F}$. The three axioms are as follows:

Axiom 1. For any $E \in \mathcal{F}$, $P(E) \geq 0$.

Axiom 2. $P(S) = 1$.

Axiom 3. Suppose $E_1, E_2, \ldots, E_i, \ldots$ is a countable collection of mutually exclusive sets from $\mathcal{F}$. Suppose that the union $\cup_{i=1}^{\infty} E_i$ is also in $\mathcal{F}$. Then

$$P\left\{\cup_{i=1}^{\infty} E_i\right\} = \sum_{i=1}^{\infty} P(E_i).$$

Axiom 3 is referred to as *countable additivity.* For most probability models, we assume that if $A \in \mathcal{F}$, then $A^c \in \mathcal{F}$ as well. Furthermore if the collection of sets E_i are in $\mathcal{F}$, then so is any union or intersection of any combination of these sets. We will make these assumptions below.

(a) Suppose in the following statements A,B are in $\mathcal{F}$. Prove that Axioms 1-3 imply each of the following statements.

(i) $P(\emptyset) = 0$.

(ii) $A \cap B = \emptyset$ implies $P(A \cup B) = P(A) + P(B)$.

(iii) $P(A^c) = 1 - P(A)$.

(iv) $A \subset B$ implies $P(A) \leq P(B)$.

HINT: The set $\emptyset$ is disjoint to all other sets, including $\emptyset$ itself. This means $S = S \cup \{\cup_{i=1}^{\infty} \emptyset\}$. Also note that countable additivity and finite additivity (*ie* statement (ii) above) are distinct statements.

(b) Let $E_1, E_2, \ldots$ be a countable collection of sets in $\mathcal{F}$. It is sometimes useful to construct an associated collection of sets, denoted

$$\begin{aligned} \bar{E}_1 &= E_1, \\ \bar{E}_i &= E_i \cap E_{i-1}^c \cap \ldots \cap E_1^c, \;\; i \geq 2. \end{aligned}$$

Verify the following properties of $\bar{E}_i$:

(i) $\bar{E}_i \subset E_i$ for all $i \geq 1$.

(ii) The sets $\bar{E}_1, \bar{E}_2, \ldots$ are mutually exclusive.

(iii) $\cup_{i=1}^{\infty} E_i = \cup_{i=1}^{\infty} \bar{E}_i$.

(c) Prove **Boole's Inequality**: Let $E_1, E_2, \ldots$ be a countable collection of sets. Then $P\left\{\cup_{i=1}^{\infty} E_i\right\} \leq \sum_{i=1}^{\infty} P(E_i)$.

(d) Suppose world records for a given sport are compiled annually. For convenience, label the years $i = 1, 2, \ldots$ with $i = 1$ being the first year records are kept. Define the events

$$E_i = \{\text{ World record broken in year } i \;\}, \;\; i \geq 1.$$

Then, let Q_i be the probability that from year i onwards, the current world record is never broken. Prove that if $\sum_{i=1}^{\infty} P(E_i) < \infty$, then $\lim_{i \to \infty} Q_i = 1$. Verify that, in particular, if $P(E_i) \leq c/i^k$ for some finite constants $c > 0$ and $k > 1$ then $\lim_{i \to \infty} Q_i = 1$.